· TORPEDO *Ultra*: Unified Searching of Locally Mounted and Web-based Journals



NASIG 14th Annual Conference June 11, 1999

Laurie E. Stackpole Naval Research Laboratory

Putting Today's Issue in Context Status of Digital Journal Access

End users need journal information

- appearing in many journals
- published by a variety of publishers
- available from distributed Web sites
- searchable with different protocols

Libraries (and vendors) facilitate access by

- linking from journals to publisher Web sites
- linking from databases to journal articles
- providing users with a common search interface

Unified Searching of Web Journals Framework for Discussing the Issue

- The Naval Research Laboratory
 - Who we are
 - Digital Library Infrastructure
- Providing links to local and remote Web journals from

InfoWeb

Citation Database

Web OPAC

- E-mail alerting service

- TORPEDO *Ultra* for access to local journals
 - Background and overview
 - Browsing and searching
 - Advantages and challenges
- An approach to unified searching of local and remote journals
- TORPEDO *Ultra* Demonstration

The Navy's Corporate Research Capability

Naval Research Laboratory Environment

Main Campus - Washington, D.C.

3,000 Federal Employees [1,500 Scientists & Engineers]

1,000 On-site contractors

130-acre Campus 116 Buildings

Subject Focus

Physics, Chemistry, Electronics, Space Sciences

Other Locations

NRL - Stennis, MS

NRL - Monterey, CA

Parent Organization

Office of Naval Research, Arlington, VA

Paving the Way for Digital Journal Access

1989	Prototype Research Reports Imaging System
1991	InfoNet Campus-wide Information System
1992	Production Research Reports Imaging System
1993	Third-generation online library catalog - STILAS
1994	Dial-in access to the InfoNet
1995	TORPEDO for APS journals and NRL reports
1995	InfoWeb Information System and Gateway
1996	InfoVision/2000 Study Team recommended that
	Library provide all digital services via the WWW
	Increased NRL Web access to digital journals

http://infoweb.nrl.navy.mil

InfoWeb Information System and Gateway



Single point of entry to

- Library info and services
- Databases and publications
- Reviewed Web sites
 - Computer Support
 - Government Information
 - Internet Directories
 - Science Resources
- TORPEDO *Ultra* Digital
 Library Initiative

Links to Journals from InfoWeb

Electronic Journals on Publisher Web Sites

- Journals listed alphabetically
- Clicking on title connects to journal (or site)
- Journals are digital equivalent of print
 - Full text
- Within collection scope
- Journals are licensed for site use

Examples:

- Science Jane's Defence Weekly
- 300+ journals published by
 - American Chemical Society
 - Assoc. of Computing Machinery
 - Institute of Physics

- Optical Soc. of America
- SIAM
- Springer-Verlag

Links to Journals from Library's OPAC

- Hyperlinks from STILAS WebCat journal records to 220 locally mounted journals
 - Persistent URLs have been added to 856 field
 - Link connects to journal in TORPEDO Ultra

No links to remote journals YET

- Uncertain nature of "free" access
- Instability of access even with signed agreement
- Cannot always link to specific journal

• First Step Beyond Linking to Web Journals
Using Databases for Retrieving Full Text

- Databases are the established way to search across the journals of many publishers
- Traditionally they serve as pointers to information, e.g. provide a journal citation
- Adding URLs to bibliographic records enables users to retrieve full text article
- Approach favored by many journal aggregators, database vendors, libraries

Science Citation Index Expanded

• "Web of Science" Science Citation Index Expanded database mounted locally at NRL

Science

Citation Index

Expanded

- Jointly licensed by four members of National Research Library Alliance
- 1985 to present available to all
- Back files to 1972 being loaded for selective access
- Members: NASA GSFC, NIST, NSF, NRL

Content-enabling a Commercial Database Citation Database Links to Full Articles

- NRL-Institute for Scientific Information joint project
- Goal: To enhance capabilities of both Science Citation Index Expanded and TOREPDO *Ultra*

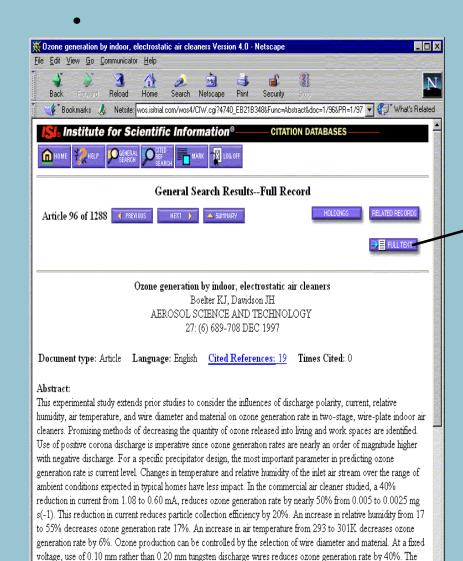
• Phase I: Add links from Citation Index to articles in TORPEDO *Ultra*

TORPEDO Ultra

Digital Archive

• Process: ISI enhances weekly database update with links to new journal data added to TORPEDO *Ultra*

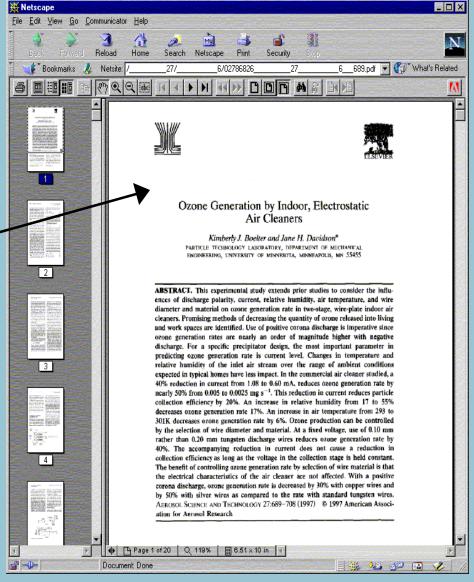
Expanded



accompanying reduction in current does not cause a reduction in collection efficiency as long as the voltage in the collection stage is held constant. The benefit of controlling ozone generation rate by selection of wire material is that the electrical characteristics of the air cleaner are not affected. With a positive corona discharge, ozone generation rate

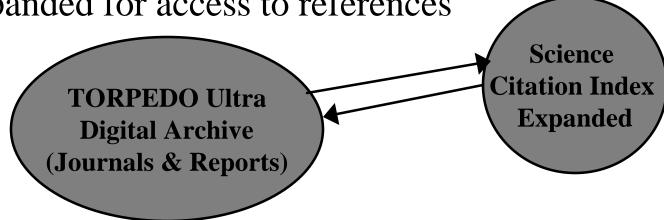
🛚 🔆 🛂 🐠 🔞 🤣

Document: Done



ISI-NRL Joint Project - Phases II and III Linking to Content of Article References

• Phase II: Add links from TORPEDO *Ultra* to article "summary" in Science Citation Index Expanded for access to references



• Phase III: Add links from Citation Index to journals residing on publisher Web sites

•

An Automated E-mail TOC Alerting Service

- About half of Library's 1,000 journals are digital
- Contents-to-Go fills in the gap
 - E-mailed journal tables of contents
 - Article requests by reply e-mail



- Researchers use InfoWeb interface to request or cancel titles, entering preferred e-mail address
- Service is fully automated
 - Vendor e-mails data to Library mail server
 - Automatically redistributed to requesters

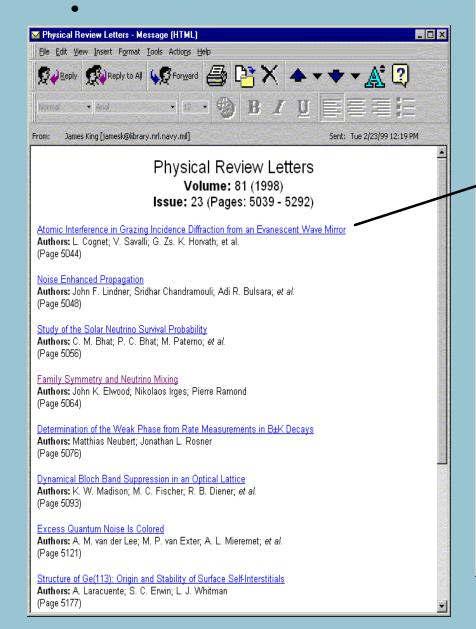
From Document *Delivery* to *Retrieval*Adding *Content* to Contents-to-Go

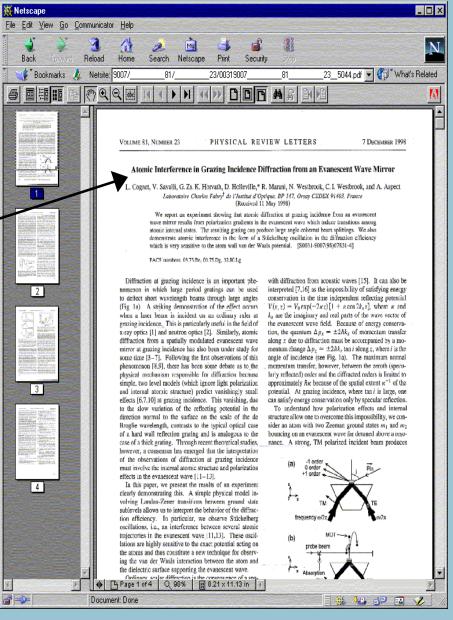
- For journals on publisher Web site, table of contents header provides URL link to <u>site</u>
- For journals in TORPEDO *Ultra*, table of contents header provides URL link to journal

• Future: tables of contents will be generated by TORPEDO Ultra and will link each title to article

Contents-to-Go
E-mail Alerts

TORPEDO Ultra
Digital Archive
(Journals & Reports)





TORPEDO Digital Library Initiative

- Desktop access to full document content
 - Equations, formulas, graphs, and charts
- Web-based user needs only network connection and graphical Web browser
- TORPEDO first went online in 1995
- Upgrade to TORPEDO *Ultra* in process

APS-NRL Journal Experiment

- 1994 Agreement between NRL and APS to test electronic dissemination of current journals *Physical Review Letters* and *E*
- Process: unbound paper copies sent by overnight mail from publisher; scanned and OCRed by Library
- Available to NRL users at their desktops within 24 hours through TORPEDO

APS-NRL CRADA- 1995

(Cooperative Research & Development Agreement)

- Investigation of search and retrieval systems for the display and printing of journal images and the use of the Internet as a transmission medium.
- By-Product: Page images of Physical Review series 1985-1995 produced by NRL
- All images are available through PROLA (Physical Review On-Line Archive)
 [PROLA URL http://prola.aps.org/]
- Most content also available through TORPEDO

TORPEDO Ultra Documents Online

- TORPEDO *Ultra* provides access to
 - 200+ Elsevier journals
 - 6 American Physical Society journals
 - 5,000 NRL Research Reports
 - 1,000 other Research Reports
 - 6,000 NRL-authored journal articles
 - 2,000 NRL Press Releases
- Coming soon
 - 17 American Institute of Physics journals

How TORPEDO Ultra Works

• TORPEDO *Ultra* offers two approaches to retrieval Journals can be **browsed** (like digital stacks)

 \rightarrow Journal \rightarrow Volume \rightarrow Issue \rightarrow Article

Or **searched** using a powerful COTS search engine (RetrievalWare by Excalibur Technologies)

• TORPEDO *Ultra* delivers text and page images TIFF-encapsulated PDF with bundled OCR Publisher-generated distilled PDF

TORPEDO *Ultra* PDF Content

- All TORPEDO *Ultra* PDF is
 - Bundled with text for searching
 - Optimized for byte serving for fast retrieval
 - Enhanced with thumbnails for document navigation
- Library gets PDF for TORPEDO *Ultra* by
 - Scanning and OCRing documents
 - Converting TIFF images to PDF
 - Obtaining PDF from publishers

RetrievalWare Search Software

- Field searching, e.g. author, title keywords, subject keywords, abstract
- Concept searching (find equivalent terms)
- Pattern searching (fuzzy-logic to find even those documents that have OCR errors)
- Boolean searching (and, or, with, near etc.)
- Search by example (find more like this one)
- Relevancy ranking

TORPEDO Ultra Hardware

Production System

- Sun 4000 Enterprise Server
 - 8 CPUs; 2GB RAM; FastEthernet
- RAID Storage
 - Level 5; 1TB Online
- Tape Jukebox Backup
 - 3.5TB capacity; 8mm Mammoth Tapes

Development System

• Sun Workstation 10

What's So Great about Local Mounting?

Advantages of TORPEDO Ultra

NRL researchers benefit through

- Local access and high speed connectivity
- A single search interface for all materials
- Integrated information from associations, commercial publishers, and government
- Ensured access to locally maintained archive
- Links from other library databases and services
- Local control and customization to meet specialized requirements

: Challenges, Prospects, and Workarounds The Library As Journal Aggregator

- Getting publishers to agree to local mounting
 - Can be time consuming (licensing, control issues)
 - May cost more (\$\$\$\$\$)
- Implementing access via a local system
 - Requires customized programming
 - May require data conversion
 - Ongoing and increasing storage requirements

: Challenges, Prospects, and Workarounds The Library As Journal Aggregator

- Many major publishers allow local mounting
 - IEEE Academic Press Kluwer
- Other publishers prohibit local mounting
 - Institute of Physics American Physical Society
- Rationale for prohibiting local mounting
 - Loss of control over copyrighted content
 - Enhancements seen as integral to digital product
 - Hyperlinks from references to cited article
 - Hyperlinks from article to subsequent citing article
 - Links to external databases and Web sites

Challenges, Prospects, and Workarounds The Library As Journal Aggregator

- So what's a Library to do?
- Objective: Make remote journals browsable and searchable along with locally mounted journals
- Potential NRL solution:
 - Add fielded and full-text indexing data for remote journals to TORPEDO Ultra
 - Retrieve article from publisher Web site by using appropriate URL as link
 - Provides end user with seamless browse and search of local and remote journals

Indexing Remote Journals: Approach #1

- Library indexing of remote journals requires publisher approval and cooperation
- American Physical Society experiment
 - Publisher delivers SGML data for indexing
 - Library
 - Extracts bibliographic data for browsing, field searching, and HTML display
 - Adds full text indexing data to TORPEDO Ultra
 - May archive SGML as backup but not for display

Indexing Remote Journals: Approach #2

- Alternative approach not yet tested
- Publisher
 - Provides bibliographic data to library
 - Allows Library to index data from Web site using a "spider"
- Library
 - Uses bibliographic data for field searching and HTML display
 - Indexes full text for TORPEDO *Ultra* searching

Conclusion

- To enable NRL researchers to search across the electronic journals of many publishers
 - Journals are mounted locally in TORPEDO *Ultra*
 - Not all required journals can be made available
- In cooperation with association publisher, NRL is testing a promising, hybrid approach
 - Index journal contents locally (for searching)
 - Link to article on publisher Web site (for display)
- Result: Integrated access to local and remote electronic journals